图书情报工作
第63卷 第4期　2019年2月

国内外社交媒体存档研究与实践述评*

黄新荣　高晨翔
西北大学公共管理学院　西安 710127

摘要：[**目的/意义**]社交媒体是网络信息资源的重要组成部分，对国内外社交媒体存档的相关研究成果进行总结与梳理，形成比较完整的综述内容，为国内该领域研究提供参考与借鉴。[**方法/过程**]基于多学科视角，通过文献计量与内容分析得到社交媒体存档领域的文献分布情况，总结社交媒体存档的理论基础，从宏观层面将社交媒体存档理论与实践研究归纳为"目的与主体""政策与指南""问题与障碍"三个研究主题并展开综述。[**结果/结论**]国内外社交媒体存档研究具有发展速度较快、与实践结合紧密、区域发展不均衡等特征，在未来研究中需要重视研究对象的变化、管理标准的构建以及多主体间的协作。最后对国内社交媒体存档研究与实践的发展提出建议。

关键词：社交媒体　网络信息资源　网络存档　数据管护　长期保存
分类号：G250.7
DOI：10.13266/j.issn.0252-3116.2019.04.015

1　引言

如今，社交媒体信息的长期保存成为了网络信息资源存档研究领域中的重要内容。2010年，美国国会图书馆(LOC)宣布与Twitter合作，将美国境内IP发布的所有推文予以存档[1]，这一极具代表性的事件有效推动了欧美发达国家社交媒体存档的实践与研究工作。同年12月，美国国家档案与文件署(NARA)发布了*Guidance on Managing Records in Web 2.0/Social Media Platforms*[2]，成为了最先颁布社交媒体数据管理政策的官方机构。2013年，美国国家数字管理联盟对美国网络信息资源存档状况进行了调查，发现受访者最为关注的存档内容即是社交媒体数据[3]。在近年的iSchool联盟年会iConference和美国档案工作者协会年会上，也均有学者提到了社交媒体存档的问题及相关的技术方案[4]。目前，来自图书情报学、档案学、计算机科学与社会学等学科的学者均从特定视角出发对社交媒体存档进行了讨论，涉及了社交媒体存档的理论基础、应用价值、问题障碍、实现技术等多方面的课题，使得社交媒体存档这一交叉领域的研究更具层次和深度。

目前，国内外关于网络信息存档的综述性文章屡见不鲜，但尚未发现有学者对社交媒体存档的研究成果进行比较全面的梳理和总结，现存的少数研究也存在学科视野单一、内容片面的现象。本文基于多学科视角对近十年以来社交媒体存档领域的研究成果进行较为系统的介绍与评述，理清相关研究的主流方向与脉络，以期为社交媒体存档的相关研究提供参考。

2　文献分布情况

2.1　文献的时间分布

本研究主要利用了Web of Science，EBSCO LISTA，EI Compendex等外文数据库以及CNKI、万方等中文数据库，经过检索、过滤得到相关英文文献225篇，中文文献128篇。对所得中英文文献进行历时分析，得出的国内外相关研究成果发表的时间分布情况见图1。结果显示，国内外学界对社交媒体存档主题进行研究均始于2000年以后。IBM实验室的L. Raymond于2001年在ACM承办的数字图书馆国际会议上发表了有关早期博客长期保存问题的文章[5]，是最早涉足这一领域的学者之一。而在中文文献方面，2006年吴开平在其硕士论文中提到了对博客等新型网络信息资源

* 本文系国家社会科学基金项目"社交媒体文件的归档与管理标准体系研究"(项目编号：16BTQ093)研究成果之一。
作者简介：黄新荣(ORCID：0000-0003-3881-8492)，公共信息资源管理系主任，副教授，博士，E-mail：huangxr@nwu.edu.cn；高晨翔(ORCID：0000-0002-6446-5822)，硕士研究生。
收稿日期：2018-06-10　修回日期：2018-08-15　本文起止页码：122-134　本文责任编辑：杜杏叶

黄新荣，高晨翔. 国内外社交媒体存档研究与实践述评[J]. 图书情报工作,2019,63(4):122－134.

进行保存的必要性和策略[6]，开启了国内相关研究的序幕。

图 1 反映出了中文文献数量在时间分布态势和特征上与英文文献之间的相似性，两者开始平稳增长的起点与峰值均存在约 6 年的时间差。结合特定的时间背景不难理解这一现象产生的原因，2004 年 Facebook 成立，带动了欧美国家社交媒体行业的发展；而 2010 年是中国的“微博元年”，国内新浪、腾讯、网易等互联网巨头都增设了微博服务，用户数量大幅增加，社交媒体在社会大众信息生活中占有很高的比重。整体而言，这一研究方向中的文献数量，尤其是国内文献的积累还比较有限，社交媒体存档仍是一个具备潜力、有待开拓的新领域。

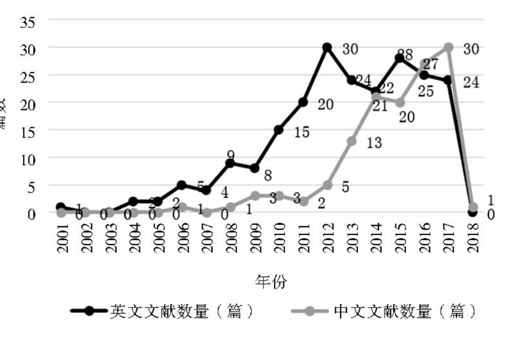


图 1　国内外社交媒体存档研究相关文献的时间分布

2.2　关键词共现分析

同一领域内高频关键词的共现关系能够反映出特定领域内重要的研究主题。由图 2 可见，英文文献中的“Social media”“Digital preservation”及“Web Archive”是出现频率最高、共现关系最复杂的关键词，后两者有效界定了社交媒体存档所属的上位研究领域。观察上述核心词汇的共现关键词，不难得出与载文分布中的相似结论——计算机科学在社交媒体存档研究中扮演重要角色，“Data mining”“Semantic Web”“Computer graphics”等也是计算机和互联网科学中的研究热点。数据挖掘和语义网技术对社交媒体存档中的信息组织与知识发现具有重要作用，计算机图形学和自然语言处理则被用于多种社交媒体数据类型的识别与分析，这些词汇的多次共现也体现了信息管理类科学与计算机科学在研究领域与技术方法上的交叉融合，社交媒体存档需要交叉学科研究以搭建起立体化、层次化的理论研究与实践工作体系。

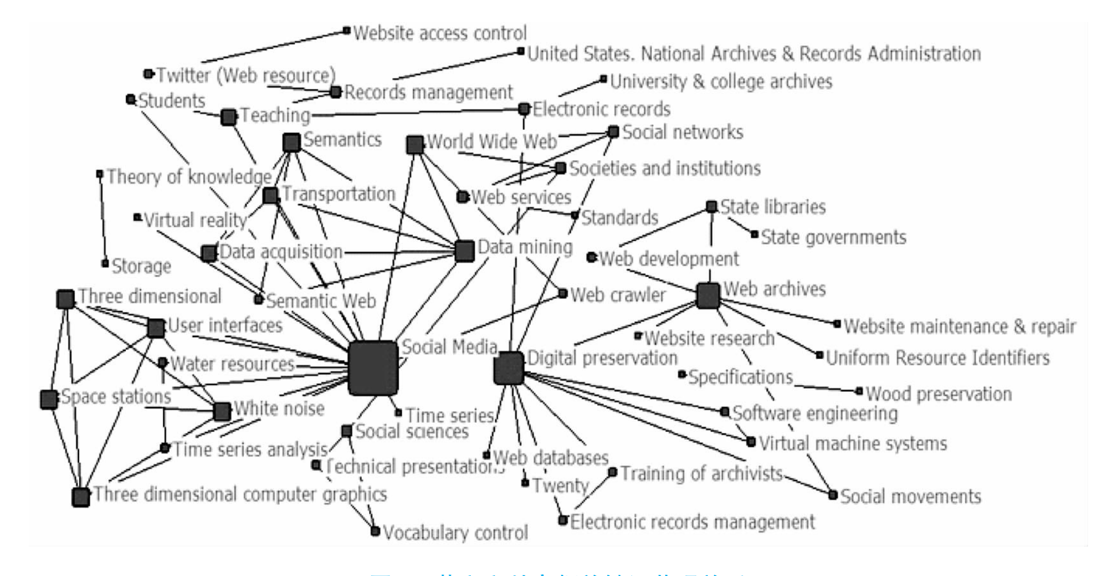


图 2　英文文献高频关键词共现关系

相比于英文文献，中文文献高频关键词共现网络的封闭性更为显著，同时体现出词性和词义的集聚现象。如“社交媒体”与“社交媒体信息”“社交媒体文件”“社交媒体档案”等名词性专业词汇以及“保存”“归档”“存档”“长期存取”等动词性的专业词汇均为高频关键词。结合词汇的期刊与作者来源，不难发现这些词汇的选取与学者的专业背景紧密相关，档案学学者更倾向于选取“社交媒体档案”“归档”，而图书情报学学者往往会选择“社交媒体信息”及“保存”“存取”等词汇。另一方面，图 3 中出现的大多为外延广泛的宏观词汇，尚未发现具体层面的策略性与技术性词汇，这也从侧面反映出目前国内社交媒体存档研究的广深度还比较有限，与计算机互联网等技术科学的融合程度有待提升。

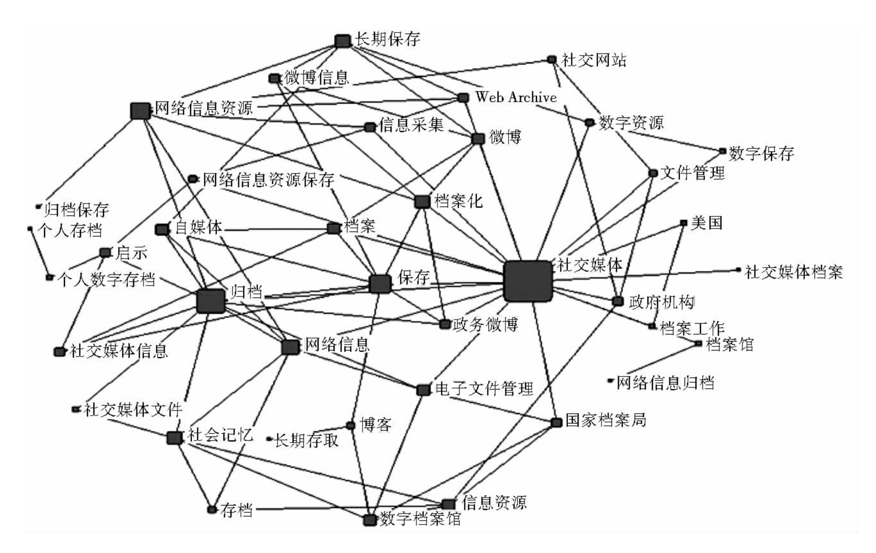


图 3　中文文献高频关键词共现关系

3　“社交媒体存档”的理论基础

3.1　“社交媒体存档”的释义与界定

“社交媒体存档”概念由社交媒体（Social media）和存档（Archiving）组合而成。Archiving 一词在 Web Archive 的相关研究中被赋予了多种定义，这些定义依其内涵可以划分为两类[7]：一类基于档案学视角，将 Archiving 翻译为“归档”，与此对应的行为是主体将具有长期保存价值的网络信息资源通过相应的技术手段保存在可控的存储介质中，注重资源的控制权和资源本身的档案特性；另一类基于信息管理视角，将 Archiving 理解为“存档”，“存档”不仅包括“归档”过程中的“采集”与“保存”环节，通常还包括资源的描述、分析、挖掘与开发等后续的信息组织与整理环节，词义及其对应的活动内容更为丰富。本文倾向于将 Archiving 理解为“存档”，认为“存档”所体现出的非线性管理思维和方式适应了社交媒体信息的保存与开发，同时也与大数据时代用户的信息需求相契合。

对“社交媒体”概念的理解直接决定社交媒体存档的基本对象和范围。社交媒体通常指“一组基于 Web 2.0 思维和技术，允许并鼓励用户生成内容（UGC）的创作及交换的互联网应用程序”[8]，其主要特征包括以下四个方面：基于 Web 2.0 模式、以用户生成内容为核心、个体或组织生成内容的设计与维护由社交媒体服务商负责、通过建立用户之间的社交关系来促进社交媒体发展[9]。结合以上定义，Facebook、Twitter、Instagram、YouTube 等英文社交媒体以及以新浪微博、腾讯微信为代表的中文社交媒体是社交媒体存档最主要的研究与工作对象，目前国内外社交媒体存档研究基本也是围绕上述几个平台展开。

3.2　社交媒体存档的相关理论

社交媒体存档以社交媒体内容为基本对象，而每一具体对象都具备特定的来源和主体，两者均是完整的社交媒体信息的重要组分。相应地，与社交媒体存档相关的理论研究同时涉及到主体及信息内容两部分，学者们既关注社交媒体信息内容作为社会记忆、数字资产的基本价值，也注重社交媒体存档中信息用户的观念和诉求。笔者对社交媒体存档的相关理论及代表性观点进行了整合，结果见表 1。

4　社交媒体存档的目的与主体

用户主体在社会生活中拥有不同层次及方向的信息需求，这决定了社交媒体存档活动也将基于不同的目的。对于社交媒体存档，其宏观目的主要包括科学研究、网络信息资源保存和司法活动。相应地，社交媒体存档的价值表现形式由其目的决定。

4.1　科研数据的保管与共享

社交媒体平台被学者认为是科研活动中的重要数据源，是科研数据获取与共享的关键渠道，对社交媒体数据进行规范的存档与管理是社交媒体数据共享的先决条件[26]。目前国内外在社交媒体科研数据的存档与共享方面尚不存在明确的行为规范与指导方案，现有的社交媒体研究数据存档与共享均为非正式行为。

4.1.1　学者团队主导的社交媒体科研数据存档　以 M. Cha 为主导的学者团队在进行 Twitter 的网络生态研究时采集并存档了一个由 17 亿条推特组成的特定数据集[27]，重点保存其中涉及的用户入度、推文转发和提及数据，以关注 Twitter 用户的影响力。该学者团队在取得 Twitter 官方许可的前提下，借助其自编的爬虫程序通过 Twitter 官方 API 对 5 200 万名用户的 Twitter 链接关系及推文内容进行采集，最终过滤得到并保存了 618 万活跃用户及其 Twitter 数据内容作为研究材料，并通过其项目官方网站“http://tinyurl.com/”提供相关 Twitter 数据共享，由于该研究目的在于识别 Twitter 用户的网络影响力，因此在 Twitter 数据存档时主要以体现用户间关系的“用户域”为数据组织单位，应 Twitter 官方的要求，该团队目前只提供特定 Twitter 数据集匿名状态下的拓扑关系，不提供具体的用户及其推文数据。

表1　社交媒体存档的相关理论概述

主要理论	作用类型/作用方式	主要观点	文献来源
社会记忆理论	基础价值 正向支持	①集体记忆建构于个体记忆之上,对个体的思维和行为习惯产生影响 ②社交媒体是社会记忆建构过程中的原始材料和资源 ③社交媒体以催化型方式介入社会记忆的维系,有利于推动权力架构趋向平衡	孙德忠[10], M. Halbwachs[11], P. Connerton[12], Library of Congress[13], F. Sushardjanti 等[14], 孙洋洋[15]
数字文化遗产理论	基础/应用价值 正向支持	①Twitter、Instagram 等社交媒体是历史记录的一部分,具备数字文化遗产容器功能,数字遗产的保存与维系提供了潜在支持 ②"非正式"的社交媒体信息在一定程度上记录了一段历史时期内的社会状态,具有作为数字资源的保存价值	C. Tiffany[4], Library of Congress[13], C. F. Tsai[16], J. Bente[17], T. Consiglio 等[18], 徐宽等[19]
信息隐私理论	应用价值 负向作用为主	①社交媒体平台直接关联用户个人,涉及用户隐私信息,易被第三方机构滥用 ②需要法律法规对社交媒体存档的基本目的和行为予以规范 ③图书馆、档案馆应介入社交媒体存档工作中来,对社交媒体信息进行统一保存与管理	B. France 等[20], R. Chakraborty 等[21], H. K. Malema 等[22], B. Xu 等[23], 万凯莉等[24], 涂海丽等[25]

日本东北大学的 M. Tatsubori[28] 团队将与2011年东日本大地震相关的社交媒体数据予以统一存档,认为这些数据是灾害管理和应急管理研究中的重要科研材料,同时设计了社交媒体科研数据存档的系统。该系统主要包括"档案类数据应用"、"外部机构合作"、"存档主平台"以及"公众数据获取"四个模块,支持其他科研机构大规模的科学数据共享请求,政府、企业、境外机构及普通民众也可通过 API 获取经平台处理过的可视化数据,其系统架构体现出简洁实用的特征。

4.1.2　科研机构主导的社交媒体科研数据存档　在科研机构方面,由德国科布伦茨——兰道大学(University of Koblenz-Landau)网络科技研究所主持的网络资源存档项目 KoNECT(Koblenz Network Collection)[29] 整合并保存了上百种不同类型的社交网络关系数据集,该研究所的科研人员将这些数据集依特定的网络类型和结构属性进行分为23个大类,并对不同的数据集进行详略有别的描述与组织,其形式包括元数据描述、社交关系方向描述、关系权重描述、关系等级描述以及添加时间戳、非完整数据集标注等,同时绘制出每一数据集的网络结构特征图,如累积度分布图、聚类系数分布图等。该项目通过其主页将社交媒体数据集向公众开放,在对数据集进行充分描述的基础上还提供原始数据的下载链接,以供其他学者进行二次分析,具有较强的开放性。

瑞士洛桑联邦理工学院支持的 CrisisLex[30] 项目则侧重于对用户发布内容的收集,该项目采集并保存了与社会危机事件相关的社交媒体数据并制作了危机数据词典。该数据集在数据的组织与描述上采用众包形式,参与者按照项目负责人提供的标签类型,依据不同数据集的主题特征对数据集进行标注。在数据利用和共享方面,CrisisLex 支持基于关键词、基于 Hashtag、基于地理位置等多样化的数据检索方式,用户可通过其主页链接到相关数据集的 GitHub 页面或直接在主页上下载 CSV 格式的原始数据。

整体而言,以上两类主体主持的社交媒体科研数据存档,其实践案例目前还比较少见,在数据的保存与共享方面仍然受到社交媒体服务商的政策制约。相比之下,仅由学者团队负责的社交媒体科研数据存档与机构负责的项目制下的社交媒体科研数据存档在保存数据的体量、数据整理的能力、存档流程的规范性、存档数据的开放性以及共享数据类型的多样性上还存在较大差距,但以上两种社交媒体科研数据存档活动能够促进其他科研团体认识到社交媒体数据的保存与管理之于科研工作的重要性。科研团体主导的社交媒体科研数据存档依赖于具体的科研项目与学者团队,相关主体能否长期、持续地对社交媒体科研数据进行存档、整理并免费向其他学者提供利用将决定这一模式的发展潜力。

4.2　网络信息资源保存

以网络信息资源保存为目的的社交媒体存档,其目标与价值因存档主体及主体认知的不同而产生了层次性的差异,具体表现为个人社交媒体数据存档、政务社交媒体存档与区域社交媒体信息资源存档三种类型。

4.2.1　个人社交媒体数据存档　社交媒体是个人表达意见及显示社会身份的渠道[31],它能够有效记录个人日常工作、生活中的重要事件;另一方面,个人社交媒体数据也是整个国家和社会信息资源的重要组分。基于以上两种认知,可以将个人社交媒体数据存档的研究视域分为微观层面的个人视角[32]和宏观层面的社会视角,不同视角下个人社交媒体数据价值的实现方式也有所不同。

微观视角下的个人社交媒体数据存档侧重于存档技术与信息系统研究，以便个人在其自身产生的零散无序的社交媒体数据集中找到所需的数据凭证或信息片段。Z. Chen 等[33]探讨了基于区块链技术（Blockchain）的个人数据存档服务系统，现已支持 Facebook、LinkedIn 等平台上的个人社交媒体数据存档。该系统基于分布式的对等网络，通过生物特性识别、个人数字组件（PDA）等模块对个人社交媒体数据内容的真实性予以认证，其中个人数字组件模块包括了对已签名数字资源进行描述的核心元数据方案，该系统的核心目的在于最大限度维护数字环境下个人社交媒体数据（尤其是反映个人成就和验证个人身份的数据）的真实性和完整性，以保障其凭证价值。D. Vianna[34]等提出的个人数字存档工具“DigitalSelf”以社交媒体数据、电子邮件和多媒体数据为主要存档对象，能够对用户产生的 Facebook、Twitter、LinkedIn 等社交媒体内容进行采集与保存。该系统主要由社交媒体采集工具、个人社交媒体数据库、社交媒体数据知识发现以及社交媒体数据检索四个模块，由采集工具通过各大社交媒体服务商提供的 API 获取社交媒体数据，并保存至社交媒体数据库，用户通过 DigitalSelf 提供的检索工具可对个人社交媒体原始数据进行检索，也可调用知识发现模块对个人数据进行进一步处理后再行检索。整体而言，DigitalSelf 是一个较为完整的个人社交媒体数据存档解决方案。

囿于个人隐私保护需求和各国信息政策法规，机构作为主体对个人社交媒体数据进行存档的实践活动较为少见，目前仅美国国会图书馆的“Twitter 存档项目”是完整且持续的个人社交媒体存档项目，以“收集美国故事并获得具有研究价值的数字资源”[35]为目的。整体而言，机构存档个人社交媒体仍处于理论探讨阶段，邱扬凡[36]认为个人或社会组织产生的非政务性微博具有信息价值、凭证参考价值以及突发事件中的社会价值，综合档案馆目前有能力也有义务对这类微博进行接收。万凯莉等[24]在界定“社交媒体用户信息”与“社交媒体个人隐私”概念的基础上，认为社交媒体服务商是一个采集存档个人社交媒体数据的可行机构，另外图书馆和档案馆作为社会记忆的保存机构也应肩负起存档个人社交媒体数据的责任。

对比二者可以发现，微观视角侧重于个人社交媒体数据的“第一价值”，即对数据生产者、形成者的基本价值，同时加入了更多个人信息隐私、数据保护方面的考量。而宏观视角侧重于个人社交媒体数据的“第二价值”，即在公开环境下对其他社会成员乃至整个社会的价值。在目前的法律与道德环境下，将个人作为存档主体更符合用户对社交媒体存档的基本认知与需求，而机构对个人产生的社交媒体数据进行存档则需要在政策法规、行为指南、技术标准等方面进行更为细致的界定，使其务必在合法合规的既定框架下进行。

4.2.2 政务社交媒体存档　在英美等发达国家，政府、国家级档案部门以及行业协会等管理组织已经明确要求业内单位应保存政务社交媒体[37]。以存档主体作为划分标准，学者对现有的政务社交媒体存档模式进行了分类，笔者分别将其命名为“双主体模式”和“三主体模式”。

“双主体模式”下的政务社交媒体存档主体包括政府机构和记忆机构两部分[38]。目前，美国和澳大利亚的中央政府直属机构均以职能部门为主体对政务社交媒体进行存档，而国家图书馆、档案馆等主要起政策制定与业务指导作用，具体实施方案由各职能部门自行决定。以 NARA 制定的政务社交媒体管理政策为指导，如海岸警卫队、住房与城市发展部、环境保护局等[39]美国政府机构都以自身职能为基础自行制定了政务社交媒体管理政策，其内容涉及了政务社交媒体的采集、保存、组织、删除等多项活动。澳大利亚的政务社交媒体存档模式与美国类似，依托其国内成熟的网络存档项目 PANDORA，澳大利亚国家图书馆对中央政府政务社交媒体存档的策略与方法进行统一规范，并由国家图书馆的内设机构指导政府各部门职员的政务社交媒体管理行为[40]，并提出了“社交媒体内容是否包含该机构政策、业务与职责的证据？这些信息是否只存在于社交媒体平台？”[41]等问题辅助政府职员判断特定政务社交媒体内容存档的必要性。

另外，记忆机构因其职能要求，也参与到政务社交媒体存档的工作中。在这方面，英国的政务社交媒体存档工作具有代表性。英国国家档案馆全权负责中央政府的政务社交媒体存档工作[42]，其工作范围仅限于内阁直属的各职能部门，而存档对象则集中在 Twitter 和 YouTube 两个政府机构常用的社交媒体平台。目前，受到英国的网络版权法案与其自身政策的严格限定，英国国家档案馆仅对英国内阁直属机构、2012 伦敦奥运会及残奥会官方平台发布的推文及相关的 YouTube 视频进行统一存档，政府部门利用 Twitter 转发与评论的数据内容、非政府部门社交媒体内容及超出网域的政务社交媒体内容均不在存档之列。

“三主体模式”是在“双主体模式”的基础上，加入

黄新荣，高晨翔. 国内外社交媒体存档研究与实践述评[J]. 图书情报工作,2019,63(4):122－134.

了第三方机构作为政务社交媒体的存档主体[43]。第三方机构作为独立的商业公司和团队，在技术水平、职能分配、关系协调等方面具有优势，能够更快地对市场需求及用户意愿做出回应。如美国社交媒体数据管理服务商 Archive Social[44] 与得克萨斯、佛蒙特、华盛顿等多个州的州政府及辖区内各市政府建立了合作关系。在社交媒体存档方面，Archive Social 主要提供两种服务方式，其一是利用公司既有的信息存储设备和信息系统对政府利用政务社交媒体产生的数据内容进行统一存档，代替政府部门对其社交媒体数据进行长期保存，并在此基础上提供了社交媒体数据分析、决策支持、知识化开发等增值服务；其二是帮助政府机关建立政务社交媒体存档政策，通过对地方政府所处内外部环境进行分析，结合风险管理与评估，为不同的政府机构制定合理的政务社交媒体管理政策，并通过案例研究的形式向政府机构作一反馈。澳大利亚的 Brolly[45] 是其国内首个与澳大利亚政府建立合作关系的社交媒体存档服务商，主要面向政府机关、企业等组织提供社交媒体存档服务。Brolly 拥有独立的云数据处理中心，能够快速实现海量数据采集、机构数据间的关联、基于数字签名的数据真实性认证等活动，该公司在存档服务方面也有着明确的使用条款和隐私政策。

以上两种模式体现了发达国家政务社交媒体存档实践的既有机制和发展水平。整体而言，“三主体模式”是“双主体模式”的发展与延续，第三方机构的介入及参与能够推动政务社交媒体存档的发展并提升政务社交媒体数据管理的规范性，这种多方合作的机制体现了社交媒体存档的发展趋势。而在研究方面，目前学者对政务社交媒体存档的研究还不够全面，政府并非政务社交媒体存档涉及的唯一主体，企业、社会组织等机构同样会面临政务社交媒体存档问题，这一类主体的社交媒体存档策略还有待学者进一步研究。

4.2.3　区域社交媒体信息资源存档　以区域为划分标准，能够使不同地区保存的社交媒体信息资源各具特色、充分反映各地自然与社会风貌。这一模式下，各地的公共图书馆、档案馆是区域社交媒体信息资源存档的核心主体，这类实践往往以各国的 Web Archive 项目为主导，将社交媒体纳入 Web Archive 的存档对象。澳大利亚国家图书馆和澳洲九所高校共同承担运营的 Web Archive 项目 PANDORA[46] 以主题事件为单位对其国内的社交媒体内容进行了存档，存档范围涉及 Facebook、Twitter 和 YouTube 等多个平台，主要由澳大利亚国家图书馆负责保存。PANDORA 在社交媒体存档时采取了保存社交媒体网页的形式，每一主题都由多个经存档的社交媒体网页构成，以此与 PANDORA 传统网络存档模式保持一致，用户可以通过事件主题、关键词、字顺等多种方式在其官网中直接检索相关的社交媒体信息。爱尔兰皇家科学院与爱尔兰国家图书馆合作，在 2014 年建立起了爱尔兰社交媒体仓储(Social Repository of Ireland)[47]，主要对全国范围内具有长期保存价值的 Twitter 内容进行存档，其后将逐步考虑更多的社交媒体平台。爱尔兰社交媒体仓储的存档流程主要包括四个步骤：事件定义、社交媒体采集与过滤、事件与推文的标注、以事件为单位进行长期保存。“事件”概念在该项目中具有重要意义，它是社交媒体信息组织的基本单位，也是社交媒体分类的标准，事件可以通过众包形式由用户输入关键词、定义标签等方式创建，随后系统会对事件主题及内容进行查重与反馈，并根据标注内容建立数据关联、剔除非法与冗余数据。

以区域层次为基础的存档形式有助于全面地构建社会记忆，是社交媒体存档实践发展最具增长潜力的形式之一。这种形式往往依托既有的网络存档项目，有地方图书馆等信息服务机构全权负责，具有一定的稳定性与可持续性。值得关注的是，区域社交媒体信息资源存档在存档对象与内容上可能会与政务机关、社会组织的社交媒体存档产生重复，这就要求地方的图书馆、档案馆等机构应对区域内的社交媒体存档工作进行统筹规划与安排，提升社交媒体存档的工作效率，避免相关数字化平台重复建设、存档主体重复工作造成的资源浪费。

4.3　数字取证

社交媒体存档对司法取证的价值主要表现为对电子证据的背景信息及元数据的全面记录。加拿大于 1998 年颁布的《统一电子证据法》(Uniform Electronic Evidence Act)中提到，对电子证据的识别性信息以及相关的变迁情况进行完整记录与存档是保障电子证据真实性的必要活动[48]。在这一方面，以信息的采集和保存为核心的社交媒体存档技术能够为司法活动中的取证环节提供支持。

M. Huber 和 M. Mulazzani[49] 仿照 Facebook 平台提供的 Neoprint，开发了一种名为“Social Snapshot”的在线社交媒体用户快照工具，该程序主要包括客户端、自动 Web 浏览器、第三方社交媒体快照程序、网络嗅探器、数字图像取证和数据分析六个模块，旨在通过多线程环境下的社交媒体数据爬取与分析形成以图像为核心的数字证据。在数字图像采集之外，Social Snap-

shot 能够结合爬虫工具对在线社交媒体用户的 ID、好友列表等特定背景信息及其他隐藏信息进行存档，将特定用户行为生成时间戳（timestamp）并以时间线（timeline）的形式与图像结合保存，经过存档打包后作为完整的电子证据。

J. Bushey[50]基于 InterPARES 项目的研究成果，讨论了社交媒体环境下用户生成内容（UGC）尤其是图像和视频存档的完整性和可信性问题，认为社交媒体信息成为长期可信数字证据的前提是解决以下问题：社交媒体环境下用于图像存档的设备必须支持 EXIF 格式（Exchangeable Image File）要求且其时设备应无异常；存档主体应参照 IPTC 核心及扩展元数据集（IPTC Core & Extensions metadata）以保证所存档的视频图像元数据完整性等。存档主体应及时与数字信息保存机构（包括图书馆、档案馆与可信第三方）联系，使存档后的图像视频处于更安全、专业的保管环境。

目前，基于数字取证的社交媒体存档技术形成了以爬虫技术、图像与自然语言处理技术和加密技术为核心的技术体系。相比科研数据保存和网络信息资源保存导向，数字取证导向更加注重社交媒体信息的真实性和可信性，需要网络环境下的动态信息转换为固化的电子证据。当前该领域的相关研究侧重于技术环境的搭建与测试，对社交媒体环境下数字取证政策与标准的研究还比较缺乏，需要对既有的数字取证法案、标准与技术体系进行进一步更新。

在以上三种社交媒体存档的基本导向之中，网络信息资源保存导向尤其是政务社交媒体存档目前在实践水平与研究热度方面均领先于其他类型的实践，是目前国内外学者在社交媒体存档研究领域中最为关注的子课题，产生了大量供学者研究分析的实践案例。社交媒体环境下的科研数据保存与数字取证还需要在研究与实践中与科学数据管理、电子文件管理等研究领域相结合，借助既有研究领域的成熟方法与体系来完善科研数据保存、数字取证领域的社交媒体存档策略与技术。

5 社交媒体存档的政策与指南

目前国内外与社交媒体存档政策及指南相关的研究主要包括两种类型，其一是对相关政策进行介绍、并在此基础上对国内的社交媒体存档提出建议的“引介—启示”型研究，其二是结合既有的社交媒体存档实践，对社交媒体存档政策的核心要素进行探讨的“深度分析—建构”型研究。

5.1 “引介—启示”型研究

“引介—启示”型研究以一系列社交媒体环境下政府信息资源管理的办法与指南为基础。最具代表性的政策如美国国家档案与文件署发布的 *Best Practices for the Capture of Social Media Records*[51]，英国国家档案馆发布的 *Operational Selection Policy*[52]以及澳大利亚国家图书馆发布的 *Social Media Policy*[53]等，这些文件中涉及社交媒体存档部分的内容大致相似，主要包括政务社交媒体的保存目的与价值、判断与识别方法、捕获与保存策略等内容。

C. Doran[54]对美国各级政府制定的社交媒体管理与保存政策的内容及特征进行了归纳总结，将相关政策依更新频率和详细程度两个指标划分至四个象限，如第一象限代表“内容详尽且更新频率高的政策”，第三象限表示“内容粗略且更新频率低的政策”。C. Doran 认为落在某一象限并不反映政策本身的优劣，粗略的政策也有其宏观层面的作用，执行机构在面对粗略政策时需要进一步制定具体化的附加政策，而第一象限的政策则需要不断更新以持续指导相关实践。

王焕[55]以存档主体和政策主题为主线，对英国、美国和澳大利亚的政务社交媒体存档政策进行了较为详细的解读，认为数字连续性（Digital Continuity）原则是国外政务媒体存档依据的基本原则，随后着重对英美澳三国政务社交媒体存档的范围、规则、捕获技术与利用方式等内容进行介绍，并从宏观层面探讨了政务社交媒体存档的所有权、动力问题以及既有政策之间的兼容问题，是国内较早关注国外社交媒体存档的研究之一。

目前，国内外的“引介—启示”型研究对网络上可见的政务社交媒体存档政策进行了比较完整、准确的解读，这一方面的研究就研究内容与文章数量而言也近于饱和，以上两位学者的研究在内容与形式上具有一定代表性。然而在官方政策之外，部分第三方机构如 Archive Social、Smarsh 等也针对企事业单位、非政府组织等机构的需求制定了相应的存档政策，这些政策的规范性与合理性仍有待探讨，这也是目前社交媒体存档政策研究的可见增长点。

5.2 “深度分析—建构”型研究

社交媒体存档政策的“深度分析—建构”型研究往往建立在既有政策基础上，旨在探讨既有政策的不足之处以及新政策的构建方式。C. Marshall 和 F. Shipman[56]基于用户意愿和观点调查，对社交媒体存档的制度问题进行了探讨，着重分析了其中涉及的存档

权限、内容范围、技术保证及法律问题,指出了社交媒体存档政策与制度构建中存在的障碍,认为后续的社交媒体存档政策应在内容价值、技术手段、法律依据等方面进行更细致的规定与说明。

高晨翔和黄新荣[57]对国内社交媒体存档政策进行了讨论与构建,包括社交媒体存档政策的依据与保障、平台与主体、条件与范围、存档技术与组织形式等内容,作者以国内的微博和微信平台为基础,认为存档主体主要包括政府机构、综合性档案馆和公共图书馆,主体应按其职能定期对具有保存价值的社交媒体信息进行统一保存,并基于 OAIS 框架提出了社交媒体信息长期保存的简要框架。宋魏巍[58]基于文件管理视域,分析了国内政府社交媒体存档的法律依据和管理准则,在此基础上构建了包括政府部门、档案部门和社交媒体服务商在内的多元互动的社交媒体文件管理主体协同机制,认为以上三个主体应分别起到政策制定与监督管控、标准制定与业务指导、数据存档与技术支持作用,同时建立"存档—管理—利用"三环节联动的政务社交媒体管理流程,在保证政务社交媒体数据安全的前提下便利用户的查询利用。

国外的政策建构型研究以既有的社交媒体管理与存档实践为基础,对问题与现状产生的原因进行了深度挖掘,指出了未来政策制定与修正的方向。而国内学者的研究在体系与构想方面均比较完整,不足之处在于缺乏必要的实践基础,因此这种政策的建构与论证的合理性方面还有待商榷和验证。在方法层面,国内研究应重视相关领域的实证调查,在得到具体、规范结论的基础上有针对性地进行相关政策构建。

6 社交媒体存档的问题与障碍

6.1 国外社交媒体存档中的争议与问题

6.1.1 隐私争议 大数据时代的数据分析者能够从只言片语中分析出社交媒体信息发布者的兴趣爱好、情感倾向、政治立场等隐私性内容,进而建立起特定信息与个人身份之间的关联。C. Marshall 和 F. Shipman[59]针对 Facebook 的制度化存档问题进行了用户调查,结果在参与调查的 244 人中,有 128 人认为无论如何都不应该使 Facebook 存档制度化,占总人数的 52%,而回答"应该存档"或"应该有条件地存档"的人数相加也仅有 101 人。反对 Facebook 存档的用户普遍担忧的问题有:大规模的 Facebook 存档让用户丢失了数据控制权;被某些机构或个人恶意利用(如犯罪前科或性取向被公开披露)等。而对于 Facebook 存档持支持或部分支持态度的用户则认为:社交媒体平台上的隐私泄露问题往往微不足道;可以接受存档主体有选择地、有范围地进行存档等。

依据学者的调查分析能够看出,社交媒体存档需要在社会价值、个人隐私与信息人格塑造之间找到平衡点。需要注意,存档中的隐私争议不仅涉及个人信息泄露问题,更为重要的是存档后的信息表现与内容反映,这些内容的建构特性可能会使用户在现实与虚拟生活中遭遇不必要的人格固化,从而影响到个人的发展。

6.1.2 权属问题 在社交媒体平台中,用户是社交媒体内容的发布者,享有社交媒体内容的著作权,社交媒体服务商对用户生成内容具有实际控制权。当多个用户围绕某一主题或事件发布包括原创信息及评论在内的大量内容,形成复杂的交互关系,这种情境下权属问题的界定颇有难度[60]。

C. Marshall 和 F. Shipman[61]利用 7 级 Likert 量表对 Facebook、Twitter 以及 Skype 用户的社交媒体信息权利意识进行了测评,发现三个平台用户均认为对自己发布的社交媒体信息应具有强大的控制权,担忧制度化的社交媒体存档会侵犯个人数据权利,尤其是"能否将用户数据存档后销售给亚马逊?""能否对存档数据进行分析以便广告推送?"等几个问题都招致了被调查者的强烈反对,而对"存档公共事件""经允许后获得个人公开数据保存权利"等问题的态度则比较温和。

臧国全和赵佩端[62]分析了社交媒体科研数据存档与应用中可能存在的侵权问题。首先,科学研究活动中的理论验证、元分析等研究方法可能需要获取与既有科研成果同样的社交媒体数据集,而 Facebook 和 Twitter 的使用条款均限制了已存档数据集的重复使用,未经授权的使用即为侵权行为。其次,当某一社交媒体数据集中包含受著作权保护的内容(如图片、视频等)时,那么基于该数据集的科研成果若内嵌了这些内容则同样构成侵权。尽管存在"通过相同 ID 重新存档数据"以及"图片分析仅限于元数据层面且采用定量方法"等折中办法,但大大影响了科学研究的效率和准确性。

社交媒体的存档主体往往需要采取复杂的折衷方式才能避免侵权陷阱,因此许多国家和机构尚未大规模开展社交媒体存档工作。面对这一问题,大英图书馆(The British Library)以本国的非印本文献法定呈缴制度(Non-print Legal Deposit)为保障,有选择地存档与某一重要事件或主体相关的社交媒体内容,避免了其

中存在的著作权纠纷[63]，值得其他机构借鉴。对于科研导向的社交媒体存档，其著作权问题尚需各国立法机构、科研机构及相关利益主体相互协作以提出合适的解决方案。

6.1.3 技术限制 相比于隐私和权属问题，技术限制表现出的问题更为客观，是社交媒体存档工作中最直接的障碍。F. McCown 和 M. L. Nelson[64]在研究 Facebook 的存档技术时认为，传统的 Web Archive 技术方法在面临 Facebook 存档问题时往往以两种形式出现——基于比特的存档和基于内容的存档。前者是网页存档中最常用的方式，即以网页为单位，将某一 Facebook 页面的 HTML 文件、样式表、JavaScript 编写的脚本内容以及 Flash 文件等统一存档，以便在 Web 浏览器中查看；而后者则是将 Facebook 上的图像、文本和视频抓取至本地分别保存，然而这两种存档方式可能造成社交媒体的原格式丢失和数据交互性缺失。

S. D. Thomson 和 W. Kilbride[65]则指出了社交媒体存档多个方面的技术障碍，内容主要包括：

(1)访问限制。与普通网页不同，Twitter 与 Facebook 均在一定程度上禁止数据共享，限制了社交媒体数据的访问能力。

(2)采集与捕获限制。科研活动中的社交媒体存档往往需要基于 API 进行批量的数据采集，但各社交媒体平台对存档主体的数据量请求和捕获速率均有限制。

(3)加工和利用限制。存档后的海量社交媒体数据必须依用户需求进行分析与挖掘才能发挥其价值，这对用于数据加工的软硬件平台均有一定要求；而且加工后的利用方式也受社交媒体服务商的协议限制，Facebook 和 Twitter 规定用户不能随意“出售、出租、重新发布、垄断”社交媒体数据。

通过以上学者的分析论述不难发现，社交媒体存档的技术障碍通常表现在技术层面，但并非由单纯的技术原因引起。如今，云计算、语义网和人工智能技术的不断发展，爬虫程序的进一步智能化、高效化，云计算带来的信息处理能力与存储能力以及基于深度学习的文本、图像处理技术为打破社交媒体存档的技术限制提供了极大帮助，因此单纯的技术限制是暂时的。解决服务商对数据的管控与限制问题、用户对数据泄露与隐私风险的担忧问题能够从更深层次为技术问题的解决提供思路与办法。

6.2 我国社交媒体存档发展的制约因素

相比国外，我国的社交媒体存档实践发展缓慢。前文提到的国外社交媒体存档的争议与问题也是制约我国社交媒体存档发展的重要原因，此外，我国学者还总结了一些影响我国社交媒体存档发展的个性化因素，具体表现如下。

6.2.1 存档主体的意识问题 以当前社交媒体存档中的主流模式——政务社交媒体存档为例，存档主体意识的缺失对社交媒体存档工作的开展具有明显影响。目前我国从中央到地方各级各类国家机关都开设了各自的官方微博、微信，但负责运营的工作人员以兼职居多，往往不具备社交媒体信息管理意识，也普遍缺乏专业的社交媒体存档与管理能力[66]。笔者曾对我国东部地区某国家机关政务微博、微信平台的存档情况进行了网络调查，负责运营的工作人员均认为微博、微信是开放平台，过往信息的检索和查询较为容易，无需专门存档。因此，这一认识环境下的社交媒体存档工作自然难以开展。

6.2.2 政策保障与业务指导缺位 通过对国外社交媒体存档实践的总结与分析，不难发现政策法规的保障和专业机构的指导是社交媒体存档实践的必要基础，而我国目前在以上两方面均存在不同程度的缺位。就政策法规而言，我国目前在微观层面缺乏具体的社交媒体管理、保存政策，宏观层面的网络信息法律体系也尚未建立，无论是政务机关还是其他主体发布的社交媒体，对其进行存档都缺乏必要依据，这也导致社交媒体存档缺乏“财政拨款”这一重要的经费来源[25]。另一方面，前文提到的重要存档主体——图书馆、档案馆目前均未将社交媒体纳入长期保存的对象范畴，也没有设置专门机构或安排人员对社交媒体存档工作进行业务指导，因此无论政务机关或是社会组织，在社交媒体存档方面均面临无从着手的问题。

6.2.3 社交媒体服务商的限制 我国社交媒体服务商制定的社交媒体信息管理方案也存在一定问题。据学者调查，新浪微博、腾讯微信等社交媒体平台公布的信息服务协议与隐私政策等集中存在以下问题[67]：主要强调服务商对信息的控制权，信息保存的期限、范围与技术方案均由服务平台决定，社交媒体信息的发布与存档主体难以掌握这些内容；缺乏信息管理的专业性，仅从用户角度明确用户的信息权利与义务，忽视了政务信息等专门性信息保存与管理的专业流程，也未说明服务商信息管理所依据的行业标准与法规等。这些问题反映了服务商在社交媒体信息的保存与管理方面缺乏专业视角，也为专业的存档主体批量获取社交媒体信息设置了障碍。

7 结语

7.1 研究与实践的基本特征

本文主要分“目的与主体”“政策与指南”“问题与障碍”三个主题对国内外社交媒体存档领域的研究与实践热点进行了归纳与梳理，该领域的研究与实践表现出以下几个特点。

7.1.1 依托既有研究成果与环境，发展速度较快 社交媒体存档领域是多学科交叉的热门研究领域，受到了信息管理科学、计算机科学、社会学等多个学科学者的关注。在社交媒体存档研究之前，国内外在 Web Archive 领域已经积累了较为丰富的理论与实践成果，Web Archive 的研究方法与技术逐步成熟，这为社交媒体存档研究营造了良好的环境与基础。

7.1.2 实践工作是重要动力，与学术研究相互影响 分析社交媒体存档的发展历程不难发现，实践工作是这一领域研究发展的重要推力。当美国国会图书馆宣布进行 Twitter 存档工作，英国、美国及澳大利亚图书馆与档案部门策动政府直属机关进行社交媒体存档时，学术界对社交媒体存档的讨论还十分少见。发达国家的政府与图书档案部门在社交媒体存档中遇到的隐私、技术、权属等各种问题为学者提供了良好的研究素材，推动了该领域学术研究的不断发展。

7.1.3 国内与国外在研究与实践上均存在差距 目前，我国与国外在社交媒体存档的研究与实践上均存在较大差距。一方面，相比国外研究，我国在研究视角上比较单一，研究内容上也主要以引介国外社交媒体存档与管理政策和探讨国内社交媒体存档策略为主，不同学者的研究之间存在重复性，而采取的研究方法也以思辨性论述为主，缺少基于数据调查与分析的实证研究。另一方面，我国社交媒体存档实践也进展较慢，尚未形成规模化的社交媒体存档行为，也缺乏必要的政策与标准支持，目前还没有出现如英美澳等国政务社交媒体存档这样的典型案例。

7.2 对国内社交媒体存档研究与实践发展的建议

7.2.1 学术研究方面

（1）注重研究类型与范围的扩展，完善研究体系。当前，国内学者应重视研究视野与范围的扩展，在从事政策分析型研究之外，还应重点关注国内外的社交媒体存档系统与技术、用户对社交媒体存档的意愿与诉求、既有的网络存档技术标准在社交媒体存档中的应用等内容具体、紧扣实践的热门课题，从而在该领域形成比较完善的研究体系，这对于我国社交媒体存档实践的开展也有裨益。

（2）提升研究深度，避免表面研究。国内既有的社交媒体存档研究以分析国外既有政策为主，重视对各国政策内容的列举与政策背景的分析，研究内容缺乏新意与深度。一方面，国内学者应在政策内容与背景之上关注国外的法律制度、行政体制与社会环境特征，通过对比分析探寻国内社交媒体存档实践发展缓慢的深层次原因。另一方面，国内学者还应关注国外学者在相关领域内的研究取向、研究思路与研究方法，注重对研究对象特征与属性的深层挖掘。

7.2.2 存档实践方面

（1）政策与标准先行，确立规范的存档工作体系。纵观美国、英国的社交媒体存档实践，存档政策与标准的制定是存档工作持续化、规模化开展的必要条件。我国的国家图书馆、国家档案局等单位应在深入调研的基础上先行提出国内的社交媒体存档方案，随后进一步研究既有的信息管理与长期保存标准如何应用于社交媒体信息存档，并指导试点单位逐步开展社交媒体存档工作，以此为机关单位、社会组织的社交媒体存档实践建立基础与依据。

（2）建立多主体协作下的社交媒体存档机制。多主体协作机制是国外社交媒体存档发展的基本趋势，建立多主体协作机制对于理顺各方关系、兼顾各方利益以保证社交媒体存档工作顺利开展具有重要意义。其中图书馆、档案馆与国家机关的协作是基本形式，此外，第三方商业机构在存档技术与专业性方面具有优势，也可作为政务社交媒体存档的解决方案，以上主体同时还应与社交媒体服务商建立合作协议，以便在数据的获取与管理方面取得相关权利，协助服务商完善其平台信息管理政策，保障这一机制的正常运转。

（3）注重存档实践开展的阶段性与层次性。社交媒体存档工作是一项长期性、持续性的工作，其规划与发展应避免此前各地在建设数字图书馆与数字档案馆时产生的“追赶潮流”现象，从而造成资源浪费、项目可持续性差等问题。我国在开展社交媒体存档式应循序渐进，各地应根据机构意愿选择试点进行试验性的存档工作，以便对项目的价值、效益、问题等进行评估后调整既定策略，进而有层次地建立社交媒体存档实践体系。

参考文献：

[1] Library of Congress. Twitter donates entire Tweet archive to library of congress[EB/OL]. [2018-06-02]. http://www.loc.gov/today/pr/2010/10-081.html.

[2] NARA. Guidance on managing records in Web 2.0/social media platforms [EB/OL]. [2018-06-02]. http://www.archives.gov/records-mgmt/bulletins/2011/2011-02.html.

[3] MARSHALL C, SHIPMAN F. Who owns the social Web? [J]. Communications of the ACM, 2017,60(5):52-61.

[4] TIFFANY C. Data repositories: a home for microblog archives? [C]//Proceedings of the 2011 iConference: inspiration, integrity, and intrepidity. Settle: iSchool, 2011:655-656.

[5] RAYMOND L. Long term preservation of digital information [C]//Proceedings of the ACM international conference on digital libraries. Roanoke: ACM, 2001:346-352.

[6] 吴开平. 网络档案信息资源相关问题研究[D]. 成都:四川大学,2006.

[7] 王芳,史海燕. 国外 Web Archive 研究与实践进展[J]. 中国图书馆学报, 2013(2):36-45.

[8] ANDREAS K, MICHEAL H. Users of the world, unite! The challenges and opportunities of social media[J]. Business horizons, 2010,53 (1):61.

[9] OBAR J, WILDMAN S. Social media definition and the governance challenge: an introduction to the special issue[J]. Telecommunications policy, 2015(39):745-750.

[10] 孙德忠. 社会记忆论[M]. 武汉:湖北人民出版社, 2007.

[11] HALBWACHS M. 论集体记忆[M]. 毕然,郭金华,译. 上海:上海人民出版社, 2002.

[12] CONNERTON P. 社会如何记忆[M]. 纳日碧力戈,译. 上海:上海人民出版社, 2000.

[13] Library of Congress. The library and Twitter: an FAQ [EB/OL]. [2018-06-02]. https://blogs.loc.gov/loc/2010/04/the-library-and-twitter-an-faq/.

[14] SUSHARDIANTI F, HERYBERT S, DJOKO B. Exploring collective memory and place attachment using social media data[C]//IOP conference series: earth and environmental science. Sanya: IOP Science, 2017, 99(1):12-15.

[15] 孙洋洋. 社交媒体在社会记忆建构中的介入机制探析[J]. 档案与建设, 2015(3):4-7.

[16] TSAI C F. A review of image retrieval methods for digital cultural heritage resources[J]. Online information review, 2007,31(2):185-198.

[17] BENTE J. Instagram as cultural heritage user participation, historical documentation, and curating in museums and archives through social media[C]//Proceedings of the digital heritage 2013. Marseille: IEEEXplore, 2013:311-314.

[18] CONSIGLIO T, VEER G V D. Design for living memory[C]//35th annual conference of the European association of cognitive ergonomics: transforming the everyday. Umeå: ECCE, 2017:67-68.

[19] 徐宽,任河. 数字资源长期保存的内容价值判断依据研究[J]. 图书情报工作, 2013,57(13):72-75.

[20] FRANCE B, REBORT C. Privacy in the digital age: a review of information privacy research in information systems[J]. MIS quarterly: management information systems, 2011,35(4):1017-1041.

[21] CHAKRABORTY R, VISHIK C, RAO H R. Privacy preserving actions of older adults on social media: exploring the behavior of opting out of information sharing[J]. Decision support system, 2013,55(4):948-956.

[22] MALEMA H K, MUSAKWA W. Tweets and Facebook posts, the novelty techniques in the creation of origin-destination models[J]. The international archives of the photogrammetry, remote sensing and spatial information sciences, 2016,41: 555-562.

[23] XU B, CHANG P, WELKER C L, et al. Automatic archiving versus default deletion: what snapchat tells us about ephemerality in design[C]// Proceedings of the ACM conference on computer supported cooperative work. San Francisco: ACM, 2016: 1662-1675.

[24] 万凯莉,张照余. 个人信息视角下社交媒体公开信息保存的可行性研究[J]. 档案学研究, 2014(4):17-21.

[25] 涂海丽,唐晓波. 对社会化媒体资源长期保存的思考[J]. 图书与情报, 2014(5):112-117.

[26] WELLER K, KINDER-KURLANDA K E. A manifesto for data sharing in social media research[C]//WebSci 2016 - Proceedings of the 2016 ACM web science conference. Hannover: ACM, 2016:166-172.

[27] CHA M, HADDADI H, BENEVENTO F, et al. Measuring user influence in Twitter: the million follower fallacy[C]// Proceedings of the international AAAI conference on weblogs and social media. Washington: AAAI, 2010:163-165.

[28] TATSUBORI M, WATANABE H, SHIBAYAMA A, et al. Social Web in disaster archives[C]//Proceedings of the 21st annual conference on World Wide Web companion. Lyon: W3C, 2012:715-716.

[29] University of koblenz landau. KoNECT——The Koblenz network collection[EB/OL].[2018-06-06]. http://konect.uni-koblenz.de/.

[30] Ecole polytechnique fédérale de Lausanne. CrisisLex[EB/OL]. [2018-06-06]. http://www.crisislex.org/.

[31] YAROSH S, BONSIGNORE E, MCROBERTS S. YouthTube: youth video authorship on youtube and vine[C]//Proceedings of the ACM conference on computer supported cooperative work. New York:ACM,2016,27:1423-1437.

[32] 赵跃. 数字时代个人存档研究框架的构建——从个人存档研究的定位与视角谈起[J]. 档案学研究, 2017(2):63-68.

[33] CHEN Z,ZHU Y. Personal archive service system using blockchain technology: case study, promising and challenging[C]// IEEE international conference on AI & mobile services. Honolulu: IEEE, 2017:93-99.

[34] VIANNA D, YONG A M, XIA C L,et al. A tool for personal data extraction[C]// IEEE international conference on data engineering

workshops. Chicago: IEEE,2014:80–83.
[35] Business insider. Library of congress is archiving all of America's Tweets[EB/OL]. [2018–06–06]. http://www.businessinsider.com/library-of-congress-is-archiving-all-of-americas-tweets-2013-1.
[36] 邱扬凡. 综合档案馆接收微博信息研究[D]. 南京:南京大学, 2016.
[37] MADHAVA R. 10 Thing to know about preserving social media [J]. Information management, 2011(5):33–37.
[38] 赵跃. 挑战与应对:我国政务新媒体文件归档若干问题思考[J]. 档案学通讯, 2016(3):80–86.
[39] NARA. Guidance on managing social media records[EB/OL]. [2018–06–06]. http://www.archives.gov/records-mgmt/bulletins/2014/2014-02.html.
[40] National archives of Australia. Archived websites[EB/OL]. [2018–06–06]. https://www.nla.gov.au/what-we-collect/archived-websites.
[41] National archives of Australia. Your social media policy - what about records? [EB/OL]. [2018–06–06]. http://www.naa.gov.au/records-management/agency/digital/socialmedia/social-media-policy.aspx
[42] The national archives. information on web archiving[EB/OL]. [2018–06–06]. http://www.nationalarchives.gov.uk/webarchive/information.htm.
[43] 宋香蕾. 政务微博档案化模式研究[J]. 档案学研究, 2017(1):51–56.
[44] Archival social. Resources Government[EB/OL]. [2018–06–06]. https://archivesocial.com/government-resources/.
[45] Brolly. Archive social media engagement[EB/OL]. [2018–06–06]. https://brolly.com.au/.
[46] National Library of Australia. PANDORA Australia's Web archive [EB/OL]. [2018–06–06]. http://pandora.nla.gov.au/.
[47] HARROWER N, HERAVI B R. How to archive an event: reflections on the social repository of Ireland[J]. New review of information networking, 2015(20):104–116.
[48] Uniform Law Conference of Canada. Uniform electronic evidence act (1998) [EB/OL]. [2018–06–06]. http://www.ulcc.ca/en/older-uniform-acts/electronic-evidence/1924-electronic-evidence-act.
[49] HUBER M, MULAZZANI M, LEITHNER M, et al. Social snapshots: digital forensics for online social networks[C]// Twenty seventh computer security applications conference. Orlando: ACM, 2011:113–122.
[50] BUSHEY J. Trustworthy citizen-generated images and video on social media platforms[C]//Proceedings of the annual Hawaii international conference on system sciences. Kauai: AISeL, 2015: 1553–1564.
[51] NARA. White paper on best practices for the capture of social media records [EB/OL]. [2018–06–06]. http://www.doc88.com/p-9909704119419.html.
[52] National archives. operational selection policy[EB/OL]. [2018–06–06]. http://www.nationalarchives.gov.uk/documents/information-management/osp27.pdf.
[53] National library of Australia. Social media policy[EB/OL]. [2018–06–06]. http://www.nla.gov.au/policy-and-planning/social-media.
[54] DORAN C. Governing the social network: how U.S. Federal department and agency records management policies are addressing social media content[C] // Proceedings of the annual Hawaii international conference on system science. Maui: AISeL, 2012: 2299–2308.
[55] 王焕. 国外政务社交媒体文件归档研究[J]. 档案学研究, 2015(6):99–105.
[56] MARSHALL C, SHIPMAN F. On the institutional archiving of social media[C]//ACM/IEEE-CS joint conference on digital libraries. Washington: ACM/IEEE, 2012:1–10.
[57] 高晨翔,黄新荣. 我国社交媒体文件的归档政策研究[J]. 图书馆学研究, 2017(7):47–55.
[58] 宋魏巍. 我国政府社交媒体文件管理研究[J]. 档案学通讯, 2017(1):107–112.
[59] MARSHALL C, SHIPMAN F. An argument for archiving Facebook as a heterogeneous personal store[C]//ACM/IEEE-CS joint conference on digital libraries. London: ACM/IEEE, 2014:11–20.
[60] 黄霄羽, 钱红梅. 你的网上记忆安全吗——社交媒体文件权属问题思考[J]. 中国档案, 2014(4):68–69.
[61] MARSHALL C, SHIPMAN F. Who owns the social web? [J]. Communications of the ACM,2017, 60 (5):52–61.
[62] 臧国全, 赵佩端. 社交媒体长期保存的困扰[J]. 图书馆, 2016(9):77–82.
[63] YU H H. Archiving social media in the context of non-print legal deposit[EB/OL]. [2018–06–06]. http://library.ifla.org/999/1/107-hockxyu-en.pdf.
[64] McCOWN F, NELSON M L. What happens when Facebook is gone? [C] //Proceedings of the ACM/IEEE joint conference on digital libraries. Austin: ACM/IEEE, 2009: 251–254.
[65] THOMSON S D, KILBRIDE W. Preserving social media: the problem of access [J]. New review of information networking, 2015, 20 (1/2):261–275.
[66] 周文泓. 政务社交媒体文件管理问题与对策研究[J]. 电子政务, 2017(11):91–98.
[67] 周文泓. 基于平台视角的政务社交媒体信息管理的问题与对策研究[J]. 现代情报, 2018,38(1):136–140.

作者贡献说明:
黄新荣:拟定论文选题,设计研究框架并指导论文修改;
高晨翔:文献搜集与整理,论文撰写与修改。

Review on Research and Practice of Social Media Archiving

Huang Xinrong Gao Chenxiang
School of Public Management, Northwest University, Xi'an 710127

Abstract: [**Purpose/significance**] Social media archiving has become an important content in web archive field. By combing and summarizing the related literature, this paper aims to provide reference for future related research in this field. [**Method/process**] Based on the multi-disciplinary perspective, we obtain the literature distribution features and summarized the theoretical basis of social media archive, by the means of bibliometrics analysis and content analysis. Then, we summarized the theory and practice of social media archiving as three research topics at the macro level, including "purpose and subject", "policy and guide" and "problems and obstacles". [**Result/conclusion**] This paper puts forward some characteristics for the research of social media archiving, including rapid development, closely related to practice and unbalanced regional development. Ultimately, this paper points out that scholars and practitioners should pay attention to the change of research object, construction of archiving standards and the cooperation among multiple subjects. We also give some brief suggestions to the practice and research of social media archiving in China.

Keywords: social media Web information resource Web archive data curation long-term preservation